



# PH207x: Health in Numbers and PH278x: Human Health and Global Environmental Change: 2012-2013 Course Report

## Citation

Reich, Blair Justin Fire, Sergiy O Nesterko, Daniel Thomas Seaton, Tommy Philip Mullaney, James H. Waldo, Isaac Chuang, and Andrew Dean Ho. 2014. PH207x: Health in Numbers & PH278x: Human Health and Global Environmental Change: 2012-2013 Course Report. HarvardX Working Paper Series No. 2: 1-20.

## Published Version

<http://harvardx.harvard.edu/ph207x-health-numbers-and-ph278x-human-health-and-global-environmental-change>

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11987996>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

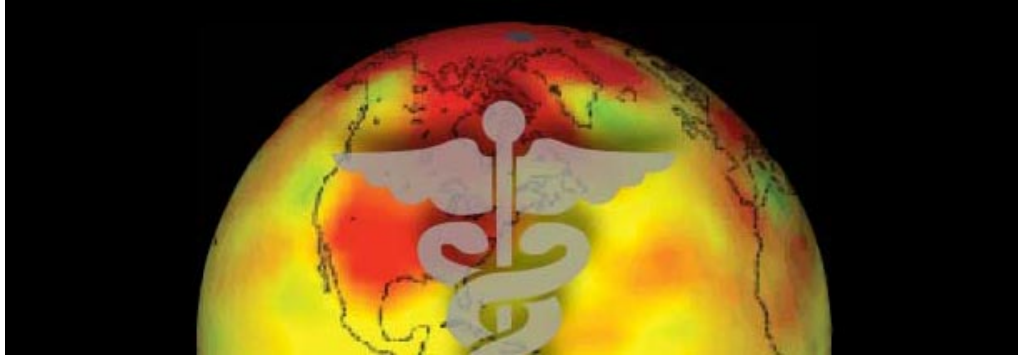
[Accessibility](#)

# PH207x: Health in Numbers & PH278x: Human Health and Global Environmental Change

## Course Report

HarvardX Working Paper #2 – 2012-2013

Winter/Spring\*



## Inside

Introduction

Structure

Students

Participation  
and Activity

Learning

The Future of  
PH207x and  
PH278x

***The School of Public Health has a mission to improve the health of the world, and one way of doing that is through education.***

**-- Professor E. Francis Cook,  
Co-Instructor PH207x: Health in Numbers**

Health in Numbers (PH207x) is the online adaptation of material from the Harvard School of Public Health's classes in epidemiology and biostatistics. The course covers the principles of biostatistics and epidemiology used for public health and clinical research.

Health and Environmental Change (PH278x) explores global environmental changes, examining their causes as well as their health consequences, and engages students in thinking about their solutions.

PH207x and PH278x were offered as HarvardX open online courses in Winter/Spring 2012-2013, on edX, a platform for massive open online courses (MOOCs). They were taught by Professors Earl Francis Cook and Marcello Pagano, and Aaron Bernstein and Jack Spengler, respectively.

**HarvardX**

**MIT** | **DL** MIT OFFICE OF  
DIGITAL LEARNING

\*Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., Chuang, I., & Ho, A. D. (2014) *Health in Numbers and Human Health and Global Environmental Change: 2012-2013 Harvard School of Public Health Course Reports* (HarvardX Working Paper No. 2).

## Introduction to the Harvard School of Public Health HarvardX Course Report

The Harvard School of Public Health (HSPH) is dedicated to the improvement of global health outcomes, in part through education. Consistent with this mission, HSPH has invested heavily in bringing significant parts of its curriculum of courses online through HarvardX. In the year ahead, HSPH aims to make courses from each of its core curriculum areas available through HarvardX and accessible worldwide.

In the 2012-2013 academic year, the first two of these courses: PH207x: Health in Numbers and PH278x: Human Health and Global Environmental Change (hereafter, “Health and Environmental Change”) were offered through HarvardX on the edX platform. This report describes the structure of these two courses, the demographic characteristics of registrants, and the activity of students. This report was prepared by researchers external to the course teams and is based on examination of the courseware, analyses of the data collected by the edX platform, and interviews and consultations with the course faculty and team members.

The report proceeds in several sections. First, we describe the goals and structure of Health in Numbers and Health and Environmental Change, with the belief that that any learning environment should be evaluated in the context of its intent, values, and vision. We then provide descriptive statistics about the students who registered for these HSPH courses and compare them to HarvardX students as a whole. With an understanding of what the course team created and metrics about the learners who took an interest in the course, we then turn to examining how participants interacted with the resources, including their patterns of assessment-taking, persistence, and overall activity. We end by examining the limits of our understanding of student learning in these courses, and describing the next steps in the development of online courses from the School of Public Health.

Our hope is that this report and its companion reports—including a [multiple-course report](#) and other reports from the first HarvardX courses—will inspire new avenues of research and provide insights to future designers of open online courses.

### The Structure of Health in Numbers

Health in Numbers, taught by Professors Earl “Fran” Cook and Marcello Pagano, was one of the first two HarvardX courses. It was offered in Fall of 2012 along with Introduction to Computer Science (CS50x). Professors Cook and Pagano volunteered early to participate in the initiative, even before it was fully clear what creating a course on edX would entail. Professor Pagano is well known to the Stata statistical programming community as one of the founders and moderators of the Stata-List. While the specific initiative of HarvardX may be new, the desire to serve the public through online learning venues predates HarvardX for nearly all of the instructors of the first HarvardX courses.

Health in Numbers takes materials from HSPH introductory courses in epidemiology and biostatistics. From the course overview:

*Quantitative Methods in Clinical and Public Health Research is the online adaptation of material from the Harvard School of Public Health's classes in epidemiology and biostatistics. Principled investigations to monitor and thus improve the health of individuals are firmly based on a sound understanding of modern quantitative methods. This involves the ability to discover patterns and extract knowledge from health data on a sample of individuals and then to infer, with measured uncertainty, the unobserved population characteristics. This course will address this need by covering the principles of biostatistics and epidemiology used for public health and clinical research. These include outcomes measurement, measures of associations between outcomes and their determinants, study design options, bias and confounding, probability and diagnostic tests, confidence intervals and hypothesis testing, power and sample size determinations, life tables and survival methods, regression methods (both linear and logistic), and sample survey techniques. Students will analyze sample data sets to acquire knowledge of appropriate computer software. By the end of the course, the successful student should have attained a sound understanding of these methods and a solid foundation for further study.*

The instructional team released content over 12 weeks, with two or three video lecture sequences released each week. Each lecture sequence consisted of five to ten videos interspersed with several quantitative or multiple choice questions about the preceding material. The first week was a general introduction to Biostatistics, and weeks 2-12 included both a biostatistics and an epidemiology sequence. Weeks 7-11 also included a third lecture sequence where a guest from the Department of Epidemiology at HSPH conducted a critique of a published paper.

The course also had several supplemental resources. During the course, students had free access to Professor Pagano’s biostatistics textbook and a free version of the Stata statistical package. Students also had access to a variety of datasets that they used to answer assessment questions and practice their statistical skills, including one dataset comprised of survey responses from students within the course.

An example of a week’s worth of work can help illuminate how these resources fit together into a course. In Week Four of Health in Numbers, the course tackled two topics: probability models for biostatistics and measures of association for epidemiology. The probability model sequence included 13 short videos, 11 by Professor Pagano and two tutorials for using Stata produced by one of the teaching fellows. Three short quizzes were interleaved among the videos, with 11 total questions about normal, Poisson, and binomial probability distributions. In addition to the videos and quizzes, the sequence also included links to probability simulation applets found online and handouts with the week’s slides, notes, and solutions to some of the worked examples. The measures of association sequence has 11 videos, all from Professor Cook, and two quizzes with five totals questions, as well as a “jotter” file where the instructors shared their lecture notes and more details about worked examples presented in the lectures.

**Table 1: Important Dates in Health in Numbers**

July 24, 2012	Registration Opens
October 15, 2012	Official Course Start Date
January 18, 2013	Final Exam Due
September 8, 2013	Date of Report Data Collection

To earn a certificate in the course, students needed to earn a total grade of 80%. Homework assignments accounted for 40% of each student's grade (with the two lowest assignment grades dropped), and the final exam accounted for 60%. The only deadline in the course was the final exam deadline of January 18, 2013. All graded material could be submitted up until that final day.

On the whole, Health in Numbers is a fairly faithful reproduction of a statistics course translated from a large lecture setting into an online setting. The course was designed with the autodidact in mind. In an interview, Professor Pagano invoked the words of Galileo: "You cannot teach anything to a man. You can only help him discover that which he has inside of him."

## The Structure of Human Health and Global Environmental Change

As with many of the first courses from HarvardX, Health and Environmental Change has a long history. The Center for Health and Global Environment at HSPH has offered versions of the course online through the Harvard Extension School and at a distance through VHS tapes since the 1990s. HarvardX represents more of an extension of a long-term vision of public-facing education rather than a new turn.

Health and Environmental Change was co-taught by Professor Jack Spengler, from the faculty of HPSH, and Professor Ari Bernstein, a pediatrician on the faculty at the Harvard Medical School. Both faculty members are deeply concerned about the challenges of anthropocentric global climate change and the public health implications of these changes. In their syllabus, they offer this introduction to the course:

*Human activity has changed the atmosphere as well as the abundance and variety of life on a global scale. Evidence is mounting that these changes may already be having serious effects on human health. This course will begin with a consideration of climate change and biodiversity as two primary examples of global environmental change and examine what they entail for human health. In the final section of the course, we will delve into some of the most promising avenues for addressing the causes of global environmental change and, in doing so, how we may ensure the healthiest possible present and future for all people.*

The course consisted of three major sections. The first two sections examined the scientific evidence related to climate change and its present and predicted impact on human health. The second section looked in particular at biodiversity, the impact of climate change on biodiversity, and the links between biodiversity and human health. In these two sections, the courseware consisted of a combination of video lectures, suggested readings, and a series of multiple choice comprehension questions. The final section of the course was a unit focused on strategies for ameliorating climate change. After a framing of this section with videos from the course faculty, the courseware consisted of a series of guest lectures from experts in climate mitigation strategies ranging from sustainable tourism to the basics of biofuels.

In the final section of the course, students participated in solution discussion groups that encouraged students to work collaboratively to research sustainable enterprises. Students joined one of 43 discussion groups, each examining a specific sector such as agricultural food

production, coffee retailers, or wood, pulp, and paper. In the first week of the exercise, students nominated an exemplar of sustainable practices from each sector. In the second week, students researched metrics by which to rate the sustainability efforts of enterprises within a particular sector (such as measuring carbon dioxide emissions in grams of CO<sub>2</sub> emitted per KM traveled for the automobile industry). In the third week, students were challenged to evaluate their nominees based on the criteria that they devised in the previous week.

**Table 2: Important Dates in Human Health and Global Environmental Change**

December 19, 2012	Registration Opens
May 15, 2013	Official Course Start Date
August 14, 2013	Final Due Date for All Graded Materials
September 8, 2013	Date of Report Data Collection

Several additional resources supplemented these core parts of the courseware. On July 28, 2013, Professors Bernstein and Spengler held a live session in Boston, hosted by Steve Curwood of PRI's *Living on Earth*; the broadcast was also live-streamed online. In addition, the faculty interviewed a handful of experts on climate change and industry leaders and made these videos available as supplementary materials in the courseware. Finally, Professor Bernstein's textbook *Sustaining Life* was suggested as a text to accompany the course.

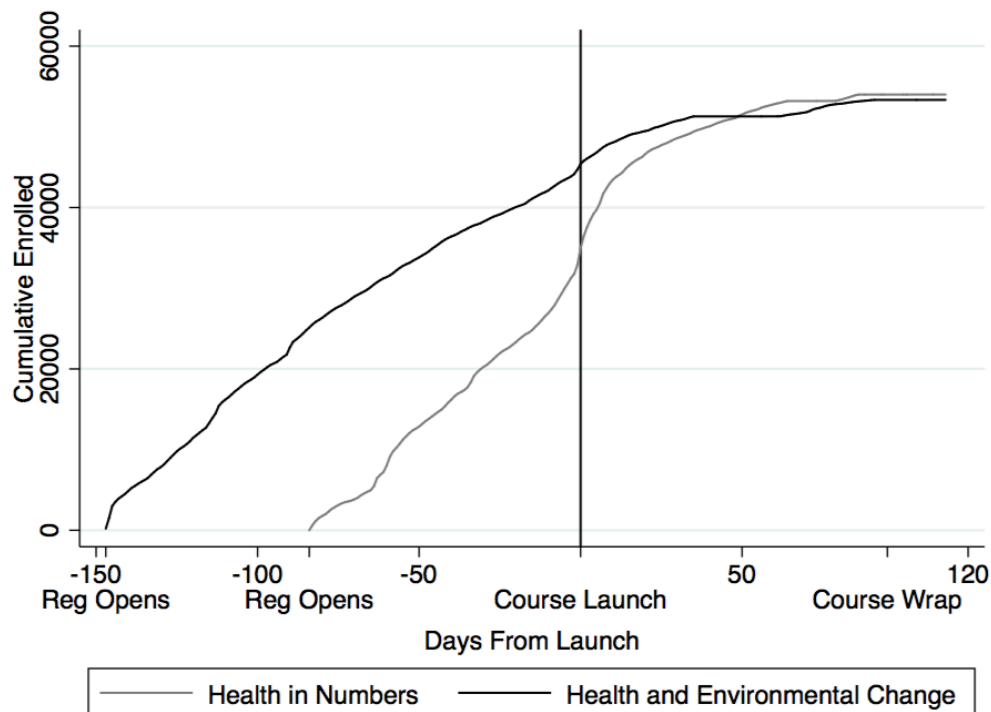
Grades in Health and Environmental Change were based on the multiple-choice questions offered throughout the first two sections. There were efforts to make contributions to the solution discussion groups a mandatory part of the course, but edX's grading and assessment system would not allow for consideration of participation in other parts of the courseware outside of the computationally-graded problems. Initially, the course team planned for grades to run from 0-100 points, with 90 points for the multiple-choice questions and 10 points for participation in the solution discussion groups, with a cutoff score of 60 for passing the course. In the final week of the course, this policy was revised so that participants with at least a grade of 50 points earned a certificate and grades ranged from 0 to 90.

### The Students of Health in Numbers and Health and Environmental Change

Registration numbers in Health and Numbers and Health and Environmental Change were similar in magnitude. As of September 8, 2013, edX had records of 61,181 students who had registered for Health in Numbers and 53,340 students who had registered for Health and Environmental Change. The difference can be attributed in part to registrations in Health in Numbers that have happened since the course "wrapped," after all graded materials were due.

Figure 1 shows the cumulative registration for Health in Numbers and Health and Environmental Change. Health and Environmental Change had a longer enrollment period before the course began by over 60 days. Health in Numbers, however, had a slightly higher registration rate during the shorter period. They had different levels of enrollment when the course launched (34,970 for Health in Numbers and 45,390 for Health and Environmental Change) and similar numbers at 120 days after course launch, just after each course wrapped (54,007 for Health in Numbers and 53,340 for Health and Environmental Change).

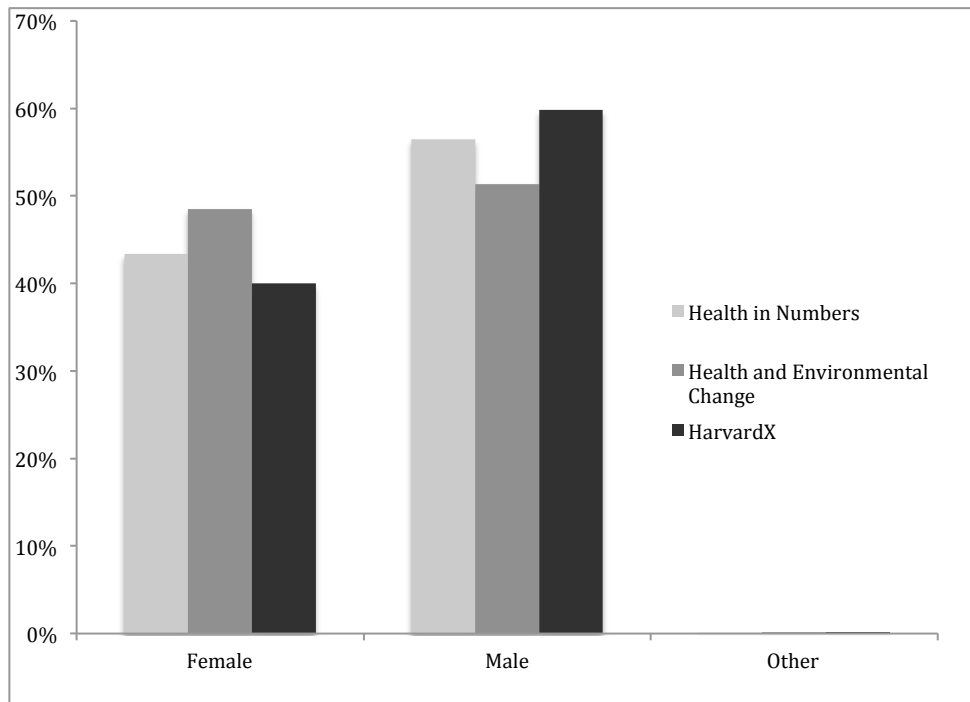




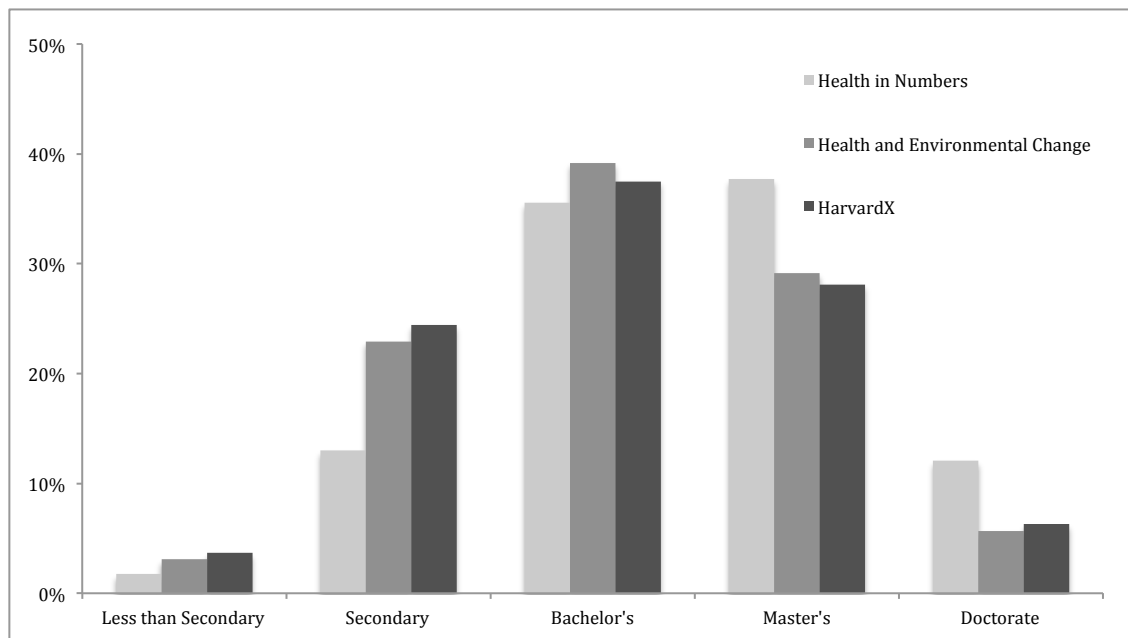
**Figure 1: Cumulative enrollment through 120 days after course launch for Health in Numbers ( $n=54,007$ ) and Health and Environmental Change ( $n=53,340$ ).**

Since Health in Numbers started nine months before Health and Environmental Change, there was an almost ten-month span between the wrap of Health and Numbers and the data collection for this report and only three weeks between the wrap of Health and Environmental Change and data collection. During this post-wrap period, over 7,000 people signed up for Health in Numbers. We include these 7,000 post-wrap registrants from Health in Numbers in many of our analyses below because they are an interesting constituency to consider. Those who register after the due date cannot have a complete course experience—they did not have access to Professor Pagano’s textbook, the free version of Stata, the discussion forums, or the final exam, and they could not earn a certificate. However, they could view the lectures, submit answers to problems and view correct answers, and take the practice exams. These students could have a meaningful, self-directed, learning experience.

In Figures 2, 3 and 4, we present data about the demographic characteristics of students in Health and Numbers and Health and Environmental Change, and we compare these characteristics to the averaged percentages from the five other initial HarvardX large scale courses (Justice, Heroes, Computer Science, Health in Numbers, and Health and Environmental Change). Health and Environmental Change was one of the more gender-balanced courses among the initial HarvardX offerings, with 49% female registrants; Health in Numbers was more typical with 43% female registrants. The [multiple course report](#) has additional details about course-specific demographics for HarvardX and MITX courses.

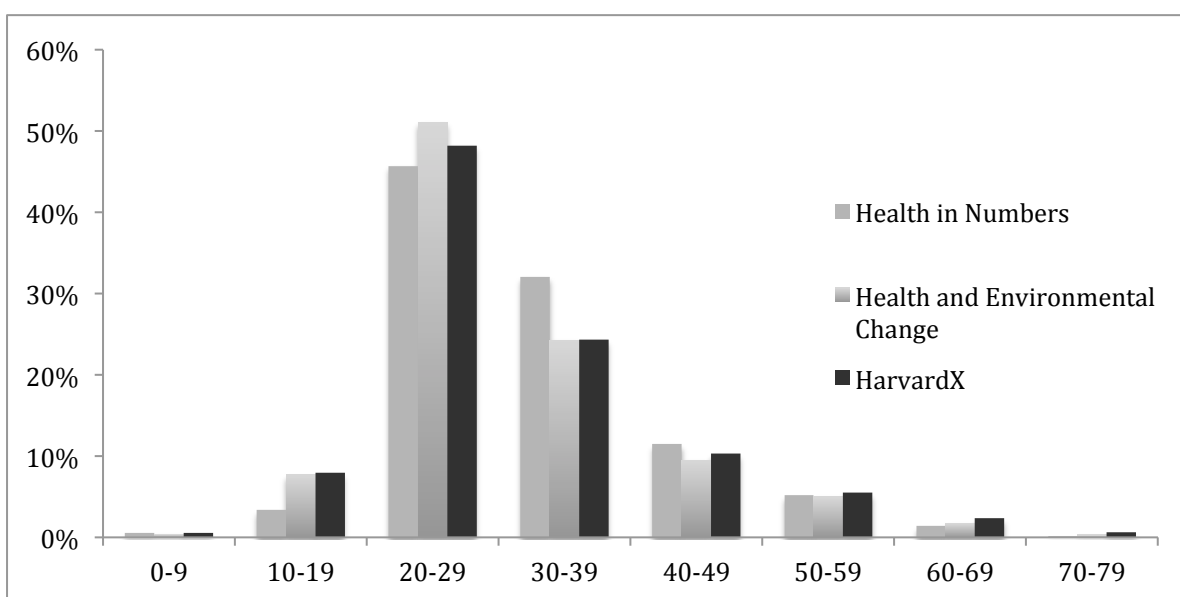


**Figure 2: Gender distribution in Health in Numbers ( $n=57,536$ ; 3,645 missing), Health and Environmental Change ( $n=50,114$ ; 3,226 missing) and the average distribution of five 2012-2013 HarvardX large-scale courses ( $n=384,060$ ; 35,254 missing).**



**Figure 3: Distribution of highest degree earned in Health and Numbers ( $n=55,638$ ; 5,5453 missing), Health and Environmental Change ( $n=48,262$ ; 5,078 missing), and five HarvardX large-scale courses ( $n=368,579$ , 50,735 missing).**





**Figure 4: Age distribution of Health in Numbers ( $n=57,000$ ; 4,181 missing), Health and Environmental Change ( $n=53,340$ ; 3,740 missing), and the average of the first five HarvardX large-scale courses ( $n=342,048$ ; 77,266 missing).**

As with other HarvardX courses, registrants in the HSPH offerings were highly educated, especially in Health in Numbers. Over 85% of students registered for Health in Numbers held at least a Bachelor's Degree. In terms of registrants' highest degree attained, 36% of registrants had a Bachelor's degree, an additional 38% had a Master's degree, and an additional 12% of students possessed doctorates. The proportion of advanced degree holders (Master's and PhD) in Health and Numbers was the highest of any of the first HarvardX courses. Health and Environmental Change was more typical of HarvardX courses: 39% of students had a Bachelor's degree, an additional 29% held a Master's degree, and an additional 6% had a PhD.

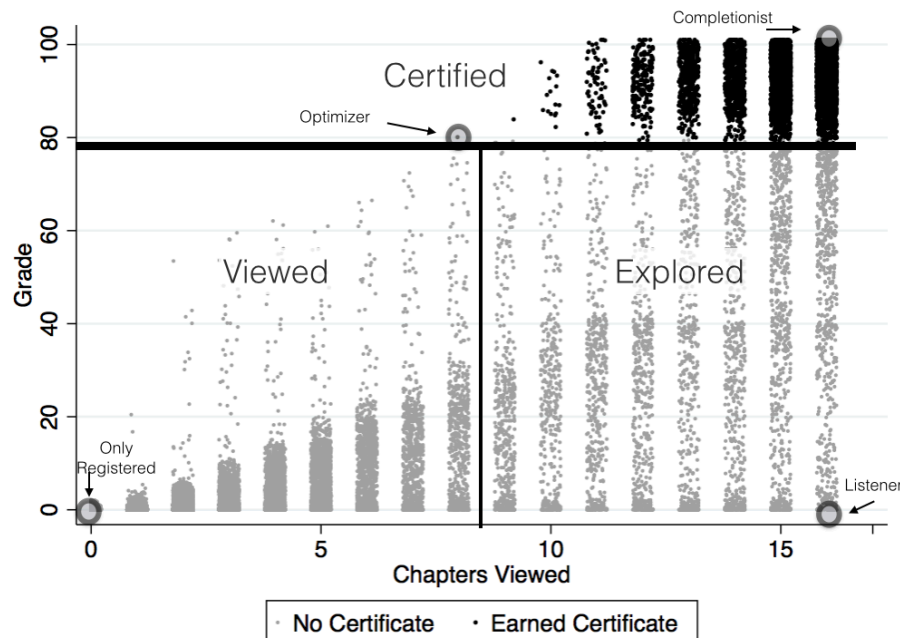
As expected, given their higher-than-average educational attainment, Health in Numbers skewed somewhat older than other HarvardX courses, with an especially high proportion of students in their 30s. Health and Environmental Change more closely tracked the average distribution of the first five large-scale HarvardX courses.

Both HSPH courses were global enterprises. Of the students with identifiable countries of residence (detected through geo-locating IP addresses and parsing self-reported addresses), 75% came from outside of the United States, with the largest second cohort in each country coming from India. The proportion of international students is higher than the other early HarvardX large-scale courses, where approximately two-thirds of identifiable registrants came from outside the United States.

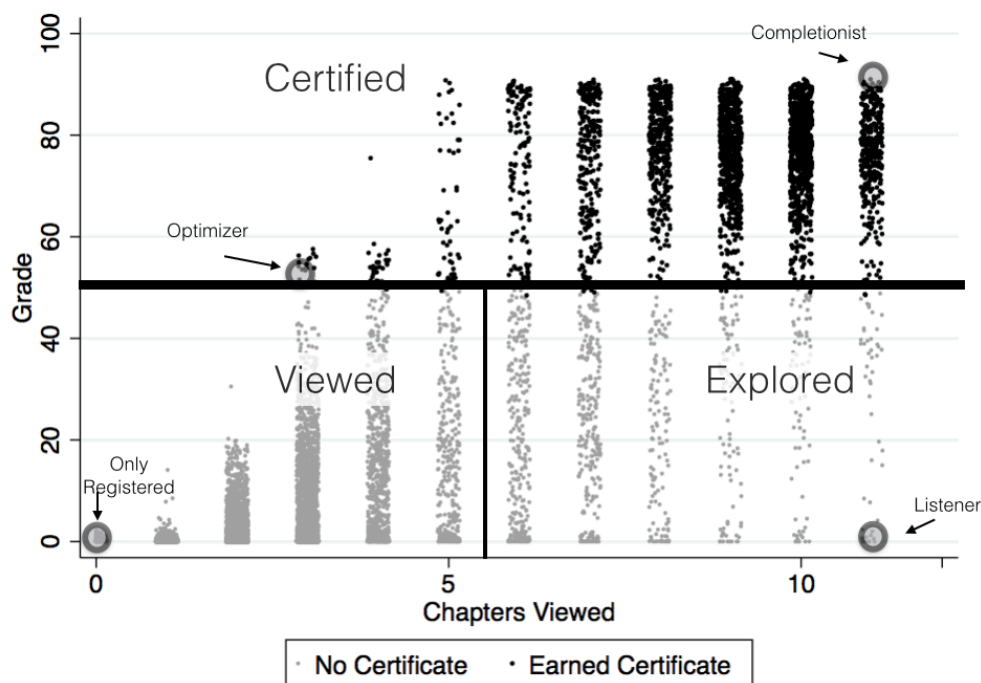
**Table 3: Country of residence for registrants with identifiable addresses in Health and Environmental Change (n=48,360; 4,980 missing) and Health in Numbers (n=58,520; 2,661 missing).**

	Health and Environmental Change				Health in Numbers	
United States	12,504	26%		United States	13,905	24%
India	4,676	10%		India	8,093	14%
United Kingdom	2,064	4%		United Kingdom	2,568	4%
Brazil	1,701	4%		Egypt	1,838	3%
Canada	1,626	3%		Brazil	1,817	3%
Nigeria	1,581	3%		Spain	1,783	3%
Germany	1,113	2%		Nigeria	1,630	3%
Spain	1,038	2%		Canada	1,385	2%
Philippines	991	2%		Australia	1,233	2%
Pakistan	857	2%		Pakistan	1,085	2%

Of the students who registered for these HSPH courses, the degree and kind of participation varied considerably. Some students completed all the materials available in the course; others focused on videos and readings while avoiding assessments; and still others focused mostly on taking assessments. To illustrate these diverse course-taking patterns, Figures 5 and 6 show scatterplots of student activity on two dimensions. On the x-axis, we plot student grades, and on the y-axis, we plot the number of “chapters” that were viewed at least once by the student (the points in the plot are jittered to show density.) Chapters are the highest-level organizational unit in the edX courseware; Health in Numbers had 16 chapters, one for each of the 12 weeks, one for a practice exam, one for the real exam, one introduction, and one collection of videos from guest lecturers.



**Figure 5: Scatterplot of grade and chapters viewed for Health in Numbers registrants (n=61,181).**



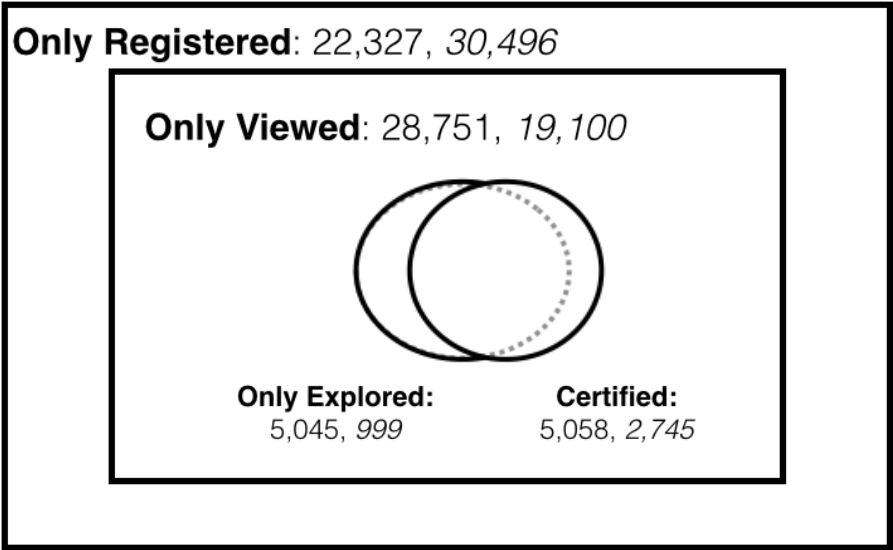
**Figure 6: Scatterplot of grade and chapters viewed for Health and Environmental Change registrants (n=53,340).**

Within these plots, we identify four interesting categories of students and several interesting specific cases. In the top of the figure, we show those students who earned a certificate in the course. In the top right, we highlight a “completionist,” a student who had the highest possible grade and also viewed all of the chapters in the course. In the top left of this top section, we highlight an “optimizer,” a student who earned a certificate with a grade exactly at the cutoff score while opening a small number of chapters in the courseware.

In the lower sections of the plots, we show students who did not earn a certificate, and we distinguish between students who viewed both more and less than half of the chapters in the course. We define those who viewed more than half of the course but did not earn a certificate as “explorers.” In the bottom right, we highlight the students who viewed all of the chapters in the courseware but answered 0 graded questions correctly as “listeners,” borrowing an MIT term for auditors. We define those who viewed less than half of the courseware and did not earn a certificate as having “viewed” the course. In the bottom left, we define those students who viewed zero chapters as “only registered.” While these points are clustered in a small space on the scatterplot, they represent a substantial portion of students in each course: 22,327 in Health and Numbers and 30,496 in Health and Environmental Change.

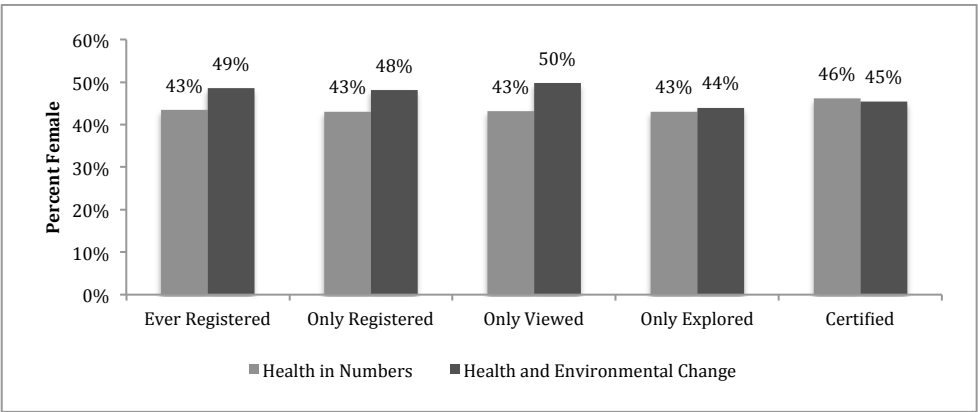
One of the signature features of these plots is that students can be found at nearly every possible location in the possibility space. Some students focused on earning a certificate by targeting assessment questions; some students viewed all parts of the course, eschewing all assessment; some students dabbled in various dimensions; and some students successfully completed all parts of the course.

Motivated by this variation (found throughout all of the initial HarvardX courses), we defined four subsamples of participants to investigate in this series of HarvardX reports: Registrants, Viewers, Explorers, and Certified. In Figure 7, we present the numbers in each group as disjoint subsets in Health in Numbers and Health and Environmental Change.



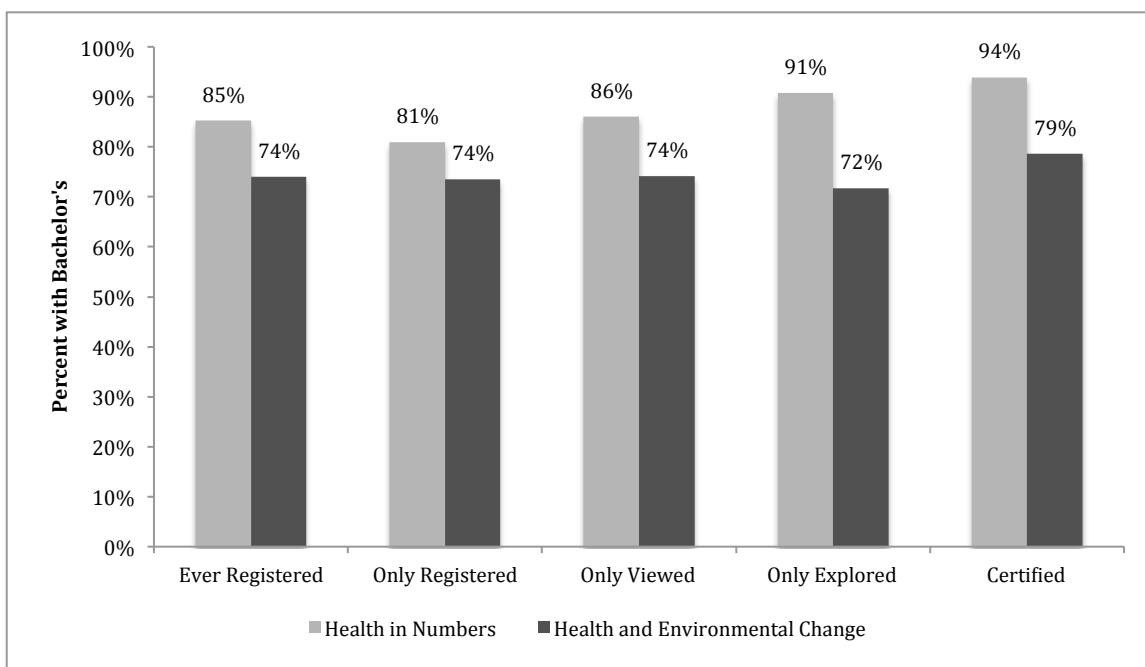
**Figure 5:** Numbers of participants in Health in Numbers and Health and Environmental Change presented in four disjoint subsets of Only Registered, Only Viewed, Only Explored, and Certified.

Examining student demographics through the lens of these categories reveals patterns of some interest. Figure 8 shows that in Health in Numbers, the female percentage was lower overall (43%) but slightly higher for certificate earners at 46%. In Health and Environment, the female percentage was higher overall (49%) but slightly lower for certificate earners at 45%.



**Figure 6:** Percentage of female students in four disjoint groups of Health in Numbers ( $n=61,181$ ) and Health and Environmental Change ( $n=53,340$ ).

We found a relationship between certificate attainment and level of education in both courses. In Figure 9, we show the distribution of the proportion of students with at least a Bachelor's degree in our four disjoint subsets, and we see that in both courses—more strongly in Health in Numbers—certificate earners were disproportionately more highly educated. Students who earned a certificate in either course also tended to be older. The median age among only registered students was 27 in Health and Environmental Change and 29 in Health in Numbers. The median age among certified students was 29 in Health and Environmental Change and 31 in Health in Numbers.



**Figure 7: Percentage of students with Bachelor's degrees in four disjoint groups of Health in Numbers ( $n=61,181$ ) and Health and Environmental Change ( $n=53,340$ ).**

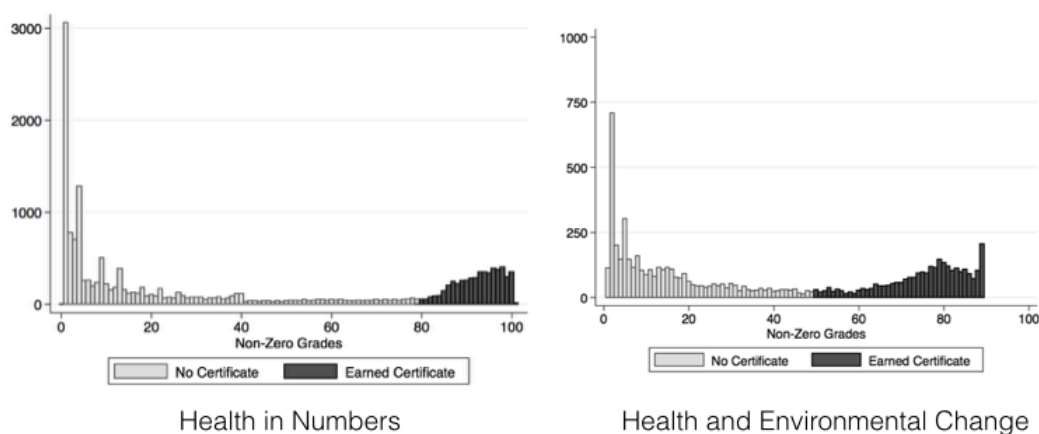
## Participation and Activity in Health in Numbers and Health and Environmental Change

### Grades and Certification

Both Health in Numbers and Health and Environmental Change used assessment questions to evaluate student competence and to award certifications of completion. Health in Numbers had multiple choice and numeric-input response questions interleaved throughout the weeks of the course, and then a final exam.

Health and Environmental Change had sets of multiple-choice questions included in the first four sections of the course. The course team attempted to use a wider range of assessments. We mentioned earlier in this report that instructors attempted to require participation in the forums for credit. Health and Environmental Change was also the first HarvardX course to pilot

machine-graded open-response questions. One question was tried early on in the course: a fairly straightforward short answer question with a correct answer. Teaching fellows from the course team graded a training set of responses, and edX attempted to grade the rest computationally, but they proved unable to differentiate between correct and incorrect responses with sufficient reliability. The question was changed to a not-for-credit question. Thus, despite efforts to expand the repertoire of assessments to include open-ended questions and participation in forums, grades in Health and Environmental Change depended entirely on multiple-choice questions. In evaluating Health and Environmental Change through the lens of assessments, it is important to recognize that grades and certificates were awarded based on the limited options that were technically feasible, not what the course team originally wanted to use to assess student competencies.



**Figure 10: Distribution of non-zero grades in Health in Numbers ( $n=17,705$ ) and Health and Environmental Change ( $n=6,722$ ).**

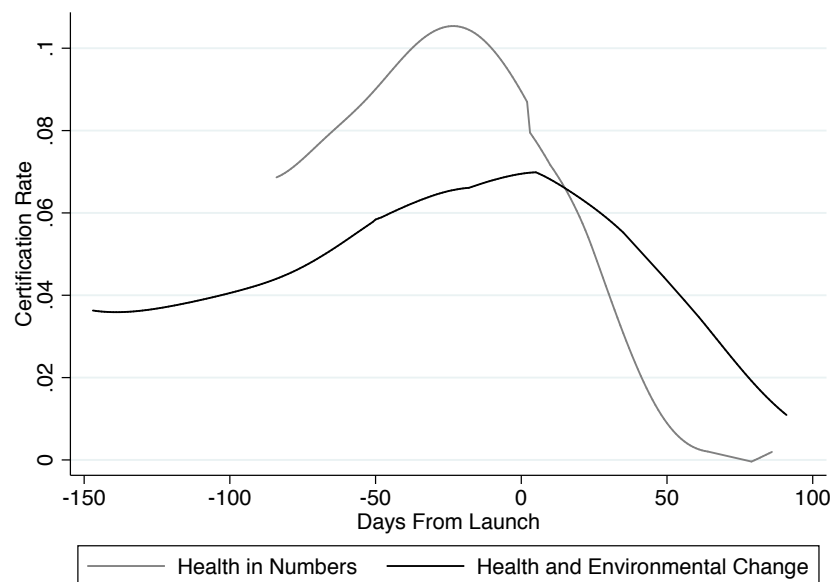
Figure 10 shows the distribution of grades in both HSPH HarvardX courses for all students who earned a grade above 0. One natural question to ask of large-scale courses is the proportion of students who earned a certificate. In most residential settings, the obvious number to use in calculating the denominator of this proportion is the number of students who enrolled in the class for credit. One would not, in a residential course for instance, responsibly calculate passing rates by including auditors in the denominator.

For the courses offered through HarvardX, we know that some enrollees register with no intent of earning a certificate, but we have no way of knowing how many. We are unable to condition our evaluation of passing rates on this critical understanding of student intent. As a set of descriptive statistics, in Table 4 we present the number of students who passed each course as a proportion of all who enrolled by the course wrap date; of all who viewed the courseware and registered before the course wrap date; and of those who opened more than half the chapters in the courseware and registered before the course wrap date. We strongly caution against drawing inferences about course quality or student learning based on these descriptive data.

**Table 4: Proportion certified of registrants, students who viewed course, and students who explored courseware in Health in Numbers and Health and Environmental Change.**

	Health in Numbers ( 5,058 certified)		Health and Environmental Change ( 2,745 certified)	
	<i>n</i>	Percent certified	<i>n</i>	Percent Certified
Registered by Wrap Date	54000	9.4%	53,337	5.1%
... and Viewed Courseware	34981	14.5%	22,843	12.0%
... and Viewed Over Half of Chapters	9,860	51.3%	3,641	75.4%

We can also examine how certification rates differ among people who register for the course at different times. Figure 11 shows an average trend line (loess approximation) of certification rate by daily registration cohort for both courses.<sup>1</sup> This simple descriptive approximation suggests that students who registered closer to the official launch of the course are more likely to earn a certificate in the course. This relationship between certification rate and registration time motivates experiments that might inform about dates to open course registration and the choice of using absolute or relative due dates.



**Figure 8: Average trend line (loess approximation) of pass rate (certified students over all registered students) by daily registration cohorts.**

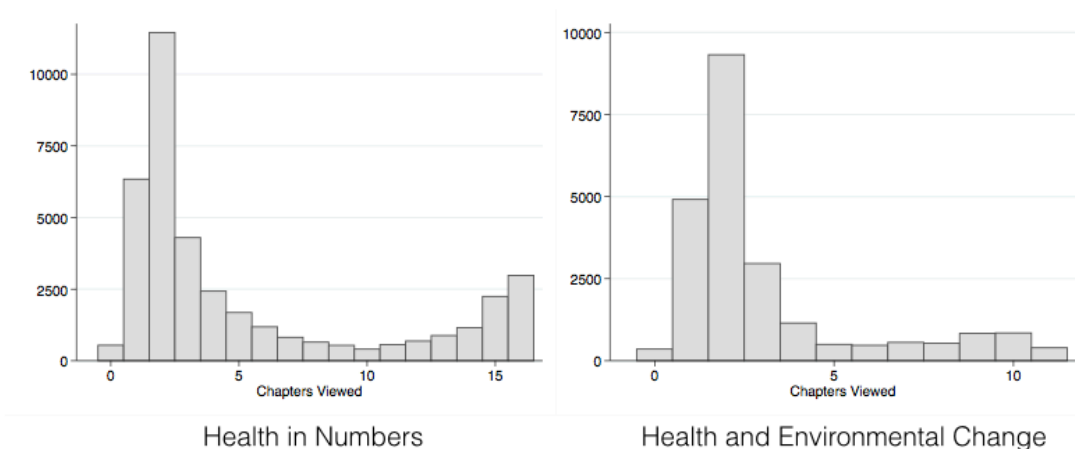
<sup>1</sup> There are some registration irregularities in both courses that have yet to be fully accounted for. In Health in Numbers, registration closed during several short periods towards the end of the course. In one day, one student was manually entered, who subsequently earned a certificate. That registration cohort was omitted.



## Activity

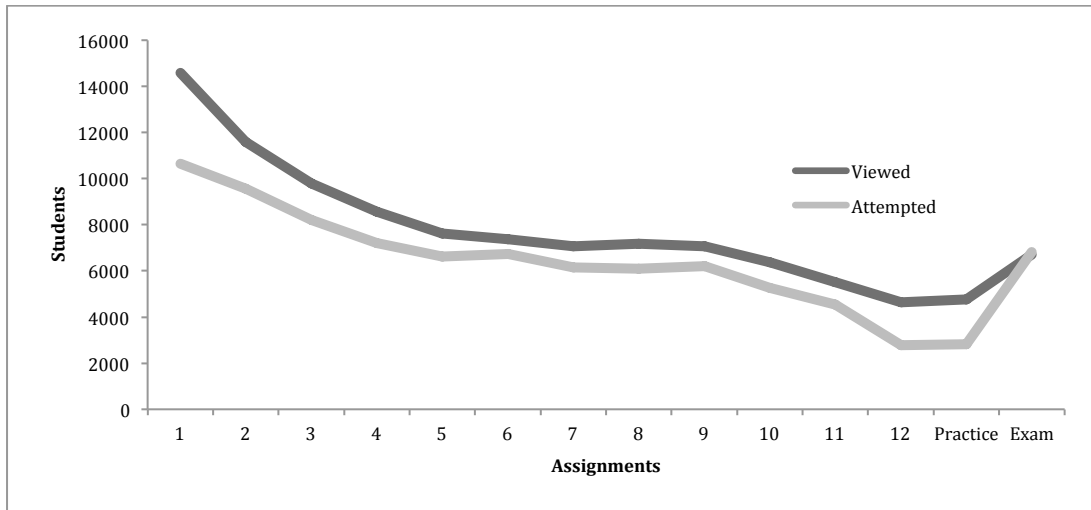
In examining student activity in the first two HSPH HarvardX courses, we present two pathways into the issue. For one perspective on student activity, we can take different pieces of courseware as the unit of analysis—for any given element of the courseware, how much did students engage with the resource? We can also take the student as the unit of analysis and ask questions about how much activity each student contributed.

A simple statistic to summarize students' activity within a course is the number of chapters of the courseware viewed by students, shown for both courses in Figure 12. This represents a metric of a necessary but substantially insufficient interaction with the course—to click on the content of a chapter one time provides no indication of how much a student might be learning from the courseware, but it is impossible to learn anything from the courseware without taking this first step. In both courses, notice the bi-modal distribution (more pronounced in Health in Numbers) where a substantial number of those who view the courseware view only a few chapters, and then a secondary mode appears around the maximum number of chapters that can be viewed. These kinds of distributions are common when looking at the usage of elements in edX courseware.

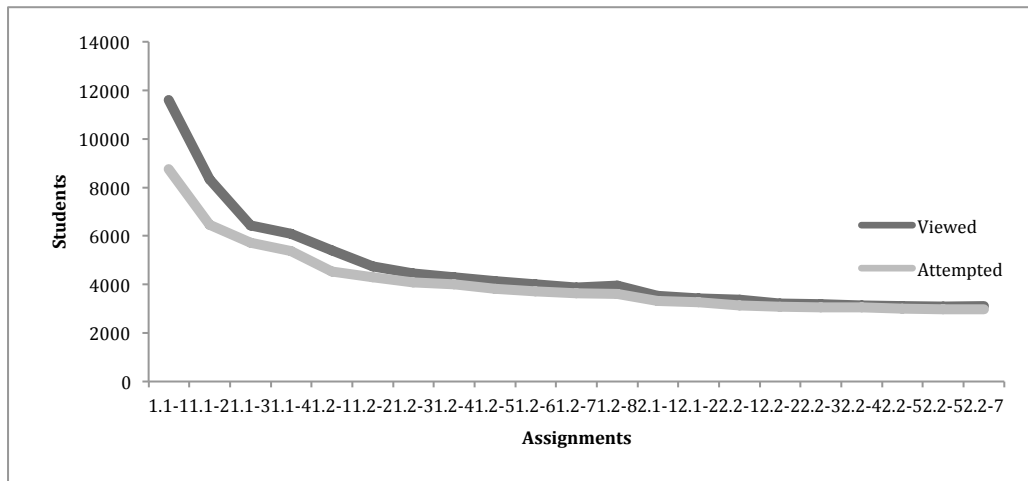


**Figure 9: Distribution of chapters viewed for participants who viewed the courseware in Health in Numbers ( $n=38,854$ ) and Health and Environment ( $n=22,844$ ).**

Another similar perspective is to look at the assessments in the course, to see how frequently they are viewed and attempted. In Figures 13 and 14, we show the average number of students who viewed and attempted the questions in each problem set for both courses. For Health in Numbers, problem views and attempts start high, decline through about the fifth week of the course, and then level off at approximately 8,000 participants viewing problems and 7,000 people attempting them. There is another drop in weeks 10 through 12. Week 12 problems were not for credit. The two lowest homework grades for each student were dropped, so students who had successfully completed the problems in weeks one through nine may have chosen not to complete problems for that reason. Of course, many other factors, including the calendar, the quality of the teaching, and the quality of assessments could also explain these findings. In the end, approximately 7,000 students took the final exam, approximately the same number of participants who were steadily answering problems throughout the course.



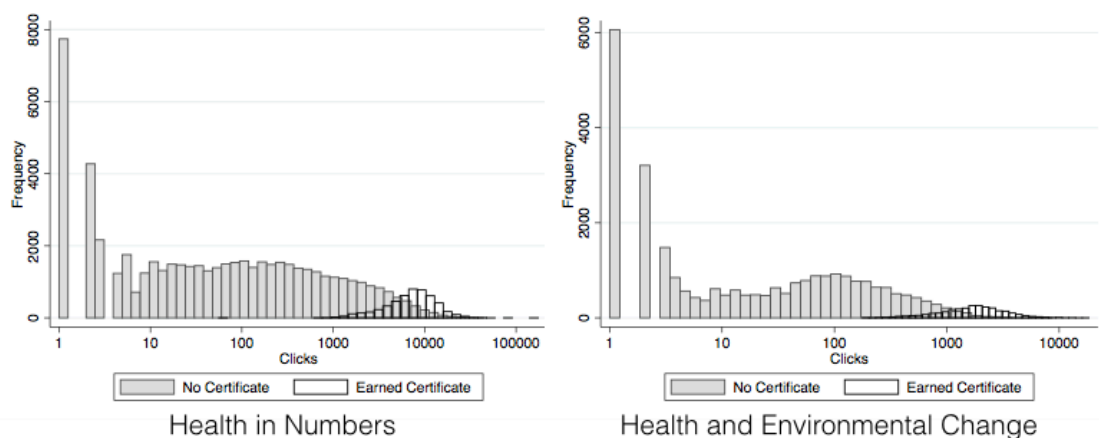
**Figure 10:** Number of students viewing and attempting problems each week for Health in Numbers.



**Figure 11:** Number of students viewing and attempting problems in each section of the Health and Environmental Change courseware.

The trends for students viewing and attempting problems in Health in Environment appear in Figure 14 and are similar. Earlier in the course, many students view and attempt problems. After the first part of the course, the decline in student activity levels off, modestly decreasing throughout the rest of the course.

Another coarse metric of student activity in a course is the number of “clicks” or actions taken by each student within the course site. Figure 15 shows the number of clicks taken by registered users in each course, for certificate earners and non-certificate earners. One simple and obvious insight is that activity predicts certification. Participants who take more actions are more likely to earn a certificate, though to be sure there are non-earners who have as many actions as certificate earners. Another ready point of comparison is that those who earned a certificate in Health in Numbers took, on average, many more actions than those who earned a certificate in Health and Environmental Change. Health in Numbers was a longer course with many more problems and videos.



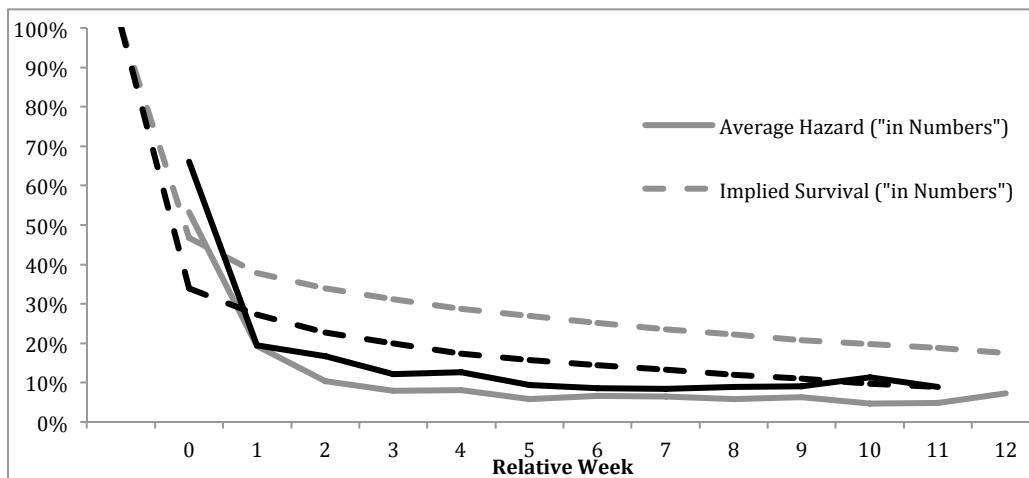
**Figure 12: Number of participant clicks (i.e. recorded actions) plotted on a log scale for Health in Numbers certificate earners ( $n=5,058$ ) and non-earners ( $n=56,136$ ) and for Health and Environmental Change certificate earners ( $n=2,745$ ) and non-earners ( $n=50,959$ ).**

## Persistence in Health in Numbers and Health and Environmental Change

Another dimension of student participation is persistence: How long do students stay with the course before stopping? Several factors make persistence complex to calculate and articulate in these HSPH large-scale courses. First, students could enter the course at any point<sup>2</sup>, and there were no deadlines, so late registrants could complete all work up to the course wrap date and still earn a certificate. Some students could take the course at the pace at which the content was released from the course launch through the final assessments. Students could also sign up in the last week of the course, and, to borrow a phrase from contemporary television viewing habits, “binge-watch” the course, completing all assignments over a weekend.

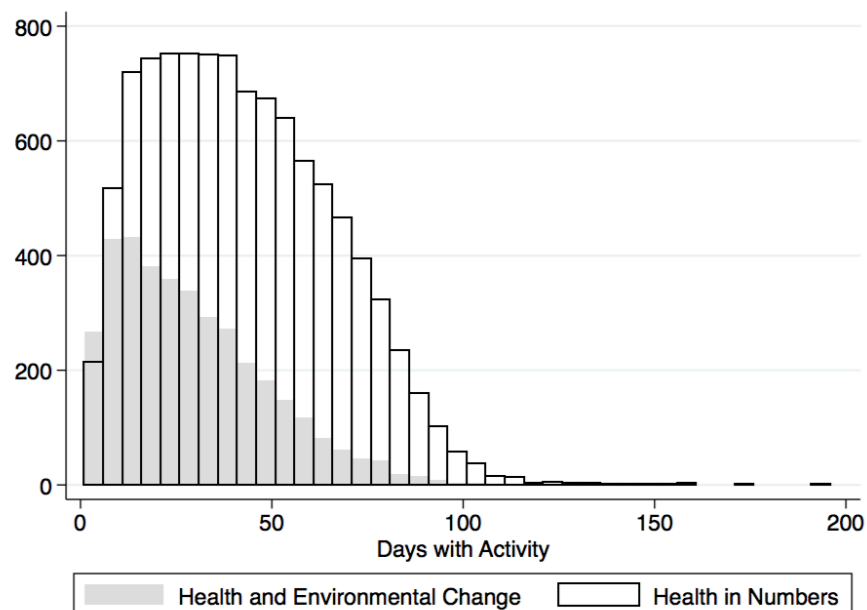
To examine student persistence, we focused more on each student’s relative timescale—the time from either their enrollment or course content launch, whichever was later—rather than on the absolute time of the course. In Figure 16, we show an average drop out probability in any given week. This function is the simple average of the drop out (or “hazard”) probabilities in each relative week. The figure answers the question, “What proportion of students who make it to their Nth week stop in that same week?”

<sup>2</sup> There were brief periods in both courses where registrations were closed.



**Figure 16:** Hazard functions comprised of the average of all weekly registration cohort hazard functions plotted on relative course week (where week 0 is the course launch or initial registration week, whichever is later). Implied survival function is calculated directly from the average hazard function.

We can then infer the proportion of students who would remain from week to week in each cohort. This is an “implied survivor function” calculated directly from the average hazard function. The key insight from this figure is that, within each weekly registration cohort, students are very likely to cease activity during the week that they register and right afterwards. After that, however, hazard proportions drop to below .2 and level out below .1 after the fifth week. Colloquially, if a student gets hooked on a course within the first two weeks—regardless of when she starts—she is likely to stay, or at least her risk of dropping out in any subsequent week is stable.



**Figure 17:** Days with activity, the number of discrete days (demarcated in UTC time) during the observational period where participants had at least one action, for explorers and certificate earners in Health in Numbers ( $n=10,102$ ) and Health and Environmental Change ( $n=3,641$ ).

This survival model gives some sense of how students persist throughout the whole course, examining the span of time from a student's entry into the course to their last action. However, a student can log in only twice, on the first day and the last day, and be counted as "surviving" through almost the length of the course. This motivates an alternative metric that counts how many discrete days students view course material. Among all registrants, we find that in Health in Numbers, the median number of days of activity is two, and 75% of registrants have nine or fewer days of activity; in Health and Environmental Change, the median number of days of activity is two, and 75% of registrants have five or fewer days of activity. For a more granular view, therefore, in Figure 17 we examine the daily activity of those who have viewed over half the course (those who "explored") and certificate earners. Among these more active learners, the median number of days active in Health in Numbers was 39 days, compared to 26 days in Health and Environmental Change. Both courses had students active on more than 100 discrete days, and Health in Numbers had a few students active on more than 150 days of the course. Since the course wrapped, in both cases, at around 90 days, this demonstrates that a few students continued to engage after the course finished.

These contrasting figures illustrate the diversity of ways of participating substantially in these HSPH courses. Most people took an action one or two days a week over the period that each course ran; however, some students did all of their participation in only a few days, and still others engaged much more frequently.

## Learning in Health in Numbers and Health and Environmental Change

Perhaps the most important question to ask of any course is: what did participants learn? In answer to this question, we can only offer cautions about our present understanding and ideas for probing this question in the future.

The limited set of assessments available to Health in Numbers and Health and Environmental Change served each course in different ways. The numeric-input responses used in Health in Numbers are appropriate for assessing many dimensions of quantitatively oriented courses. For instance, we can assess if students can correctly perform statistical tests. More nuanced challenges that face statisticians—such as evaluating the quality of interpretation or research design in a scholarly paper—are more difficult to assess using only multiple choice questions.

The assessments available on the edX platform at the time were less well suited to a course like Health and Environmental Change. The course team attempted to use a wide range of assessments, such as discussion forums and short answer questions, to assess student competency, but with limitations presented by the platform. Some of the key learning outcomes of the course, such as an invigorated commitment to address the challenges of global climate change, are hard to measure in any university setting, online or residential. Open online courses inherit from their parent institutions great challenges in understanding what students are actually learning in courses.

Moreover, we have no data about each student's level of competence coming into the course. A perfect score in the course might equally represent a student who learned all of the material for the first time as well as a student who knew everything to begin with, and wanted a refresher, or a challenge, or a certificate of completion.

With the data we have now, we have a baseline from which to track improvements in HarvardX courses, by examining ways to have more students register; to have students persist longer; or to have students engage more often in a greater percentage of the course. While these dimensions may be important, they are at best highly imperfect proxies for student learning. If we want to work toward creating more effective learning environments in a data-informed way, we need to continue to develop innovative ways to assess what students are learning and develop better understandings of what competencies and intentions students bring to a course.

## The Future of Public Health Education Online

As we write this report, two additional public health online courses from HSPH are underway and several more are in development. In partnership with faculty from the Harvard Medical School, a course on clinical trials launched on October 14, 2013, and on November 15 a Health and Society course launched examining the major social determinants of public health. Health in Numbers plans to run again in the future, and three courses are slated to launch in Spring 2014: United States Health Policy, Global Healthcare Quality and Safety, and Data Analysis for Genomics.

These online developments have led to on-campus experiments as well. HSPH students, for the first time this fall, have the option to take blended versions of the biostatistics and epidemiology introductory courses that are taking advantage of the existing courseware available in HarvardX, making more time available in class for more interactive learning. This “flipped” approach represents one way that HarvardX is encouraging experiments in online learning on campus.

As the School of Public Health moves towards having a substantial portion of its core introductory curriculum available online, there may be new innovations possible in distance and satellite learning, or blended learning on campus, and the whole may prove more than the sum of the parts. The substantial enrollments and participation in these two courses, which might be considered more of a niche subject compared to other offerings like an introduction to computer science, suggest that online courses may have a promising role to play in fulfilling HSPH’s mission of improving health outcomes and increasing quality of life around the world.